

Latent Variables, Feature Importance and Regression Session 4

Dr. Ali Arsanjani
@AliArsanjani
Adjunct At San Jose State
University
Senior Lecturer UCSD

ML in the News: social impact of ML

The Social Dilemma

PG-13 · 2020 · Documentary · 1h 33m



 Watch now
Subscription

 Watched it?

 Watchlist

[▼ All watch options](#)

8.1/10	87%	78%
IMDb	Rotten Tomatoes	Metacritic

97% liked this movie
Google users

Micro-segmentation for hyper-personalization

- Process :

1. EDAV
 1. Visualization
2. Cluster →
 1. Micro-clustering
3. (add a column with a label) → supervised learning → **Classification** →
4. select subset, categorical → numerical →
5. Regression → Prediction →
6. add context → decision
7. → action



Regression in context

Un-supervised

supervised

cluster

Logistic
regression

classify

Linear
regression

Predict

Build
context

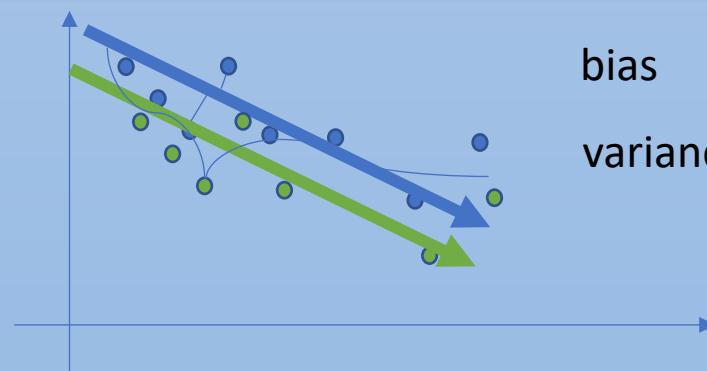
Recommend

Forecasting

(c) Ali Arsanjani, 2019

$$Y = wx + b$$

$$Y = \color{yellow}a[i] x[i] + a[0]$$



Linear Regression

Function

$$Y = f(X, w_1, b_1)$$

$$f(x, w, b) = x \cdot w + b$$

In Matrix Format

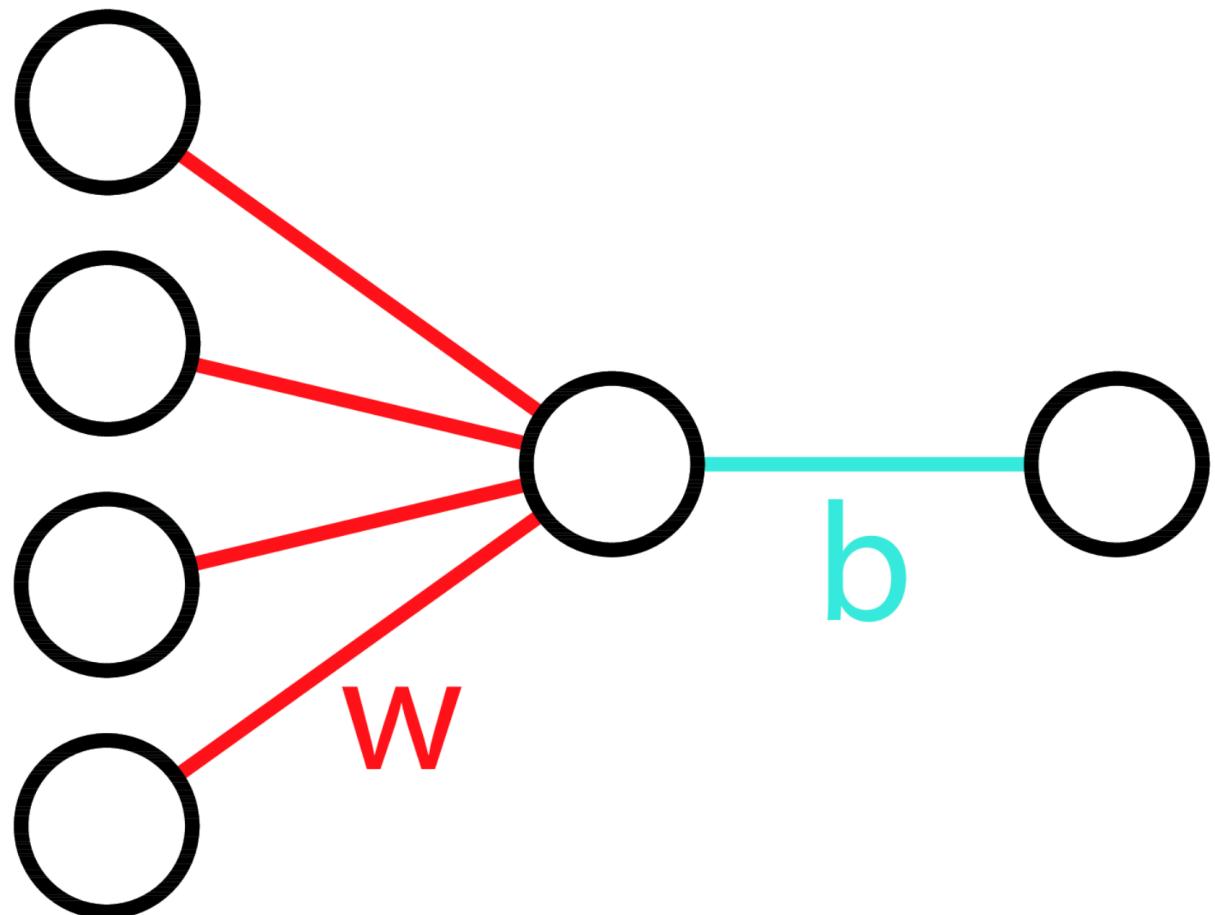
Matrices

$$X \cdot w + b = Y$$

The diagram illustrates a matrix multiplication operation. On the left, a horizontal vector X is shown as a row of four boxes, each containing three dots (...). A black dot symbol indicates multiplication by a column vector w . The vector w is represented as a vertical column of four red boxes, also each containing three dots (...). To the right of the multiplication is a plus sign (+), followed by a cyan square representing the bias vector b . An equals sign (=) follows, and finally, a square box containing three dots (...) represents the resulting vector Y .



Neural Diagram



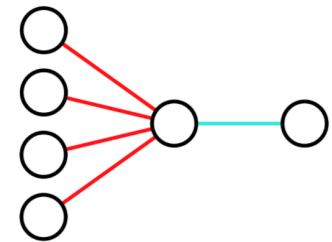
Source: [ericmjl/bayesian-deep-l](#)

LinReg 3 Ways

$$Y = f(X, w_1, b_1)$$

$$f(x, w, b) = x \cdot w + b$$

$$\begin{matrix} \dots & \dots & \dots & \dots \end{matrix} \xrightarrow{\quad X \quad} \cdot \begin{matrix} w \\ w \\ w \\ w \end{matrix} + \begin{matrix} b \\ \text{cyan square} \end{matrix} = \begin{matrix} \dots \end{matrix} \xrightarrow{\quad Y \quad}$$



Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

Logistic Regression is a binary classification solution , linear regression is a prediction (number) solution

- Logistic is a special form of linear regression
- $1/(1-e^x)$

Logistic Regression

Function

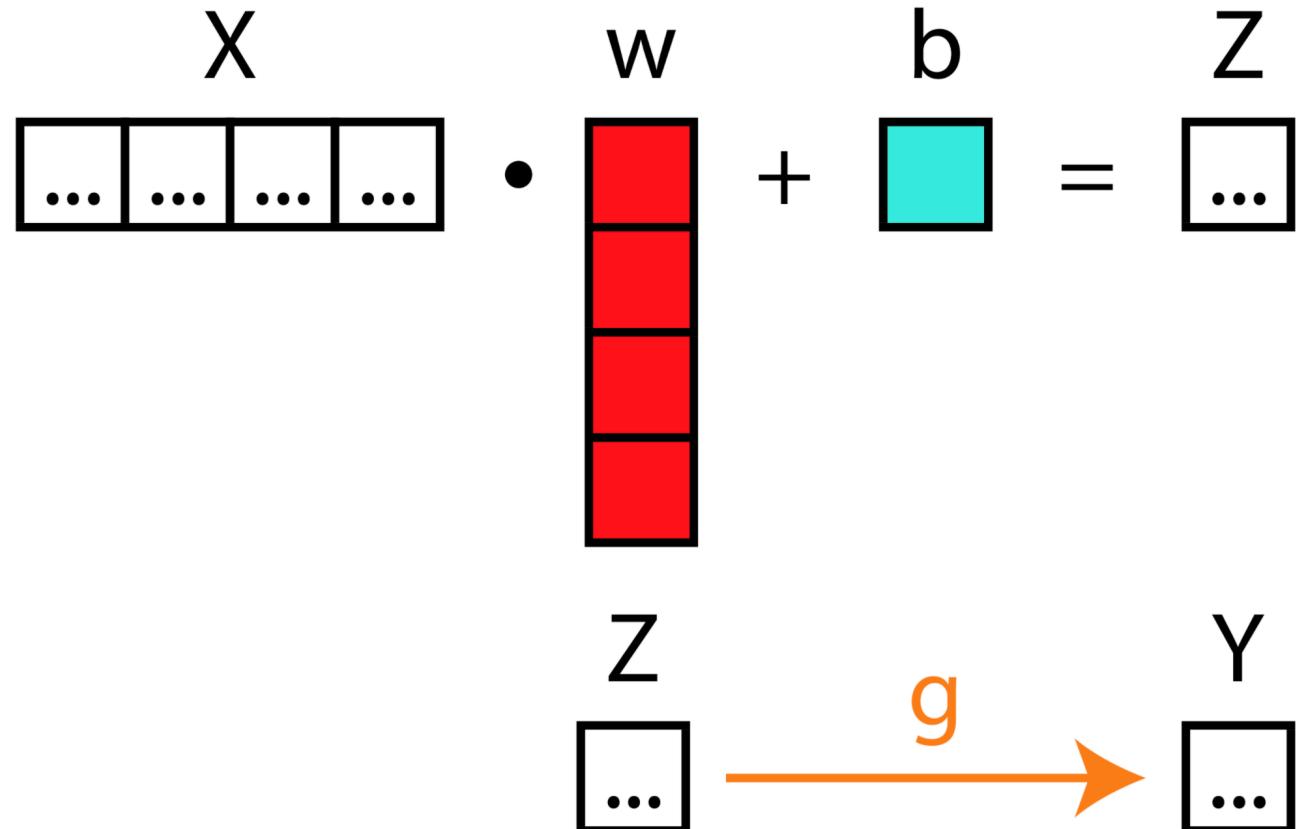
$$Z = f(X, \textcolor{red}{w}_1, \textcolor{blue}{b}_1)$$

$$Y = g(Z)$$

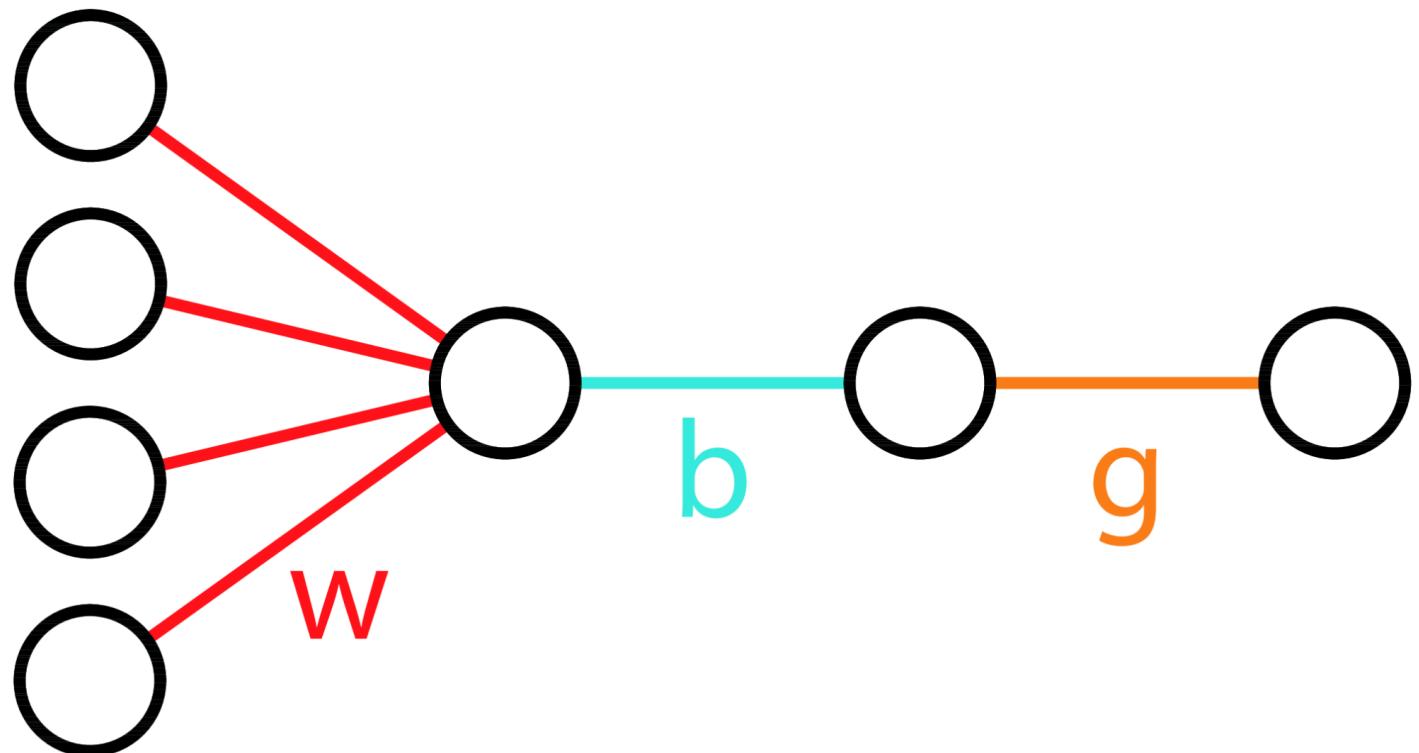
$$g(x) = \frac{1}{1 + e^{-x}}$$

Matrix View of Logistic Regression

Matrices



Neural Diagram



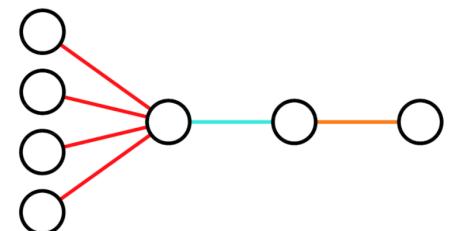
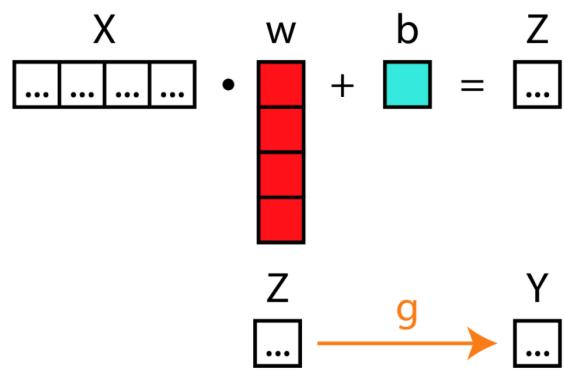
Source: ericmjl/bayesian-deep-learning-tutorial

LogReg 3 Ways

$$Z = f(X, \textcolor{red}{w_1}, \textcolor{blue}{b_1})$$

$$Y = g(Z)$$

$$g(x) = \frac{1}{1 + e^{-x}}$$



Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)



Deep Neural Networks

Function

$$Z = \tanh(f(X, w_1, b_1))$$

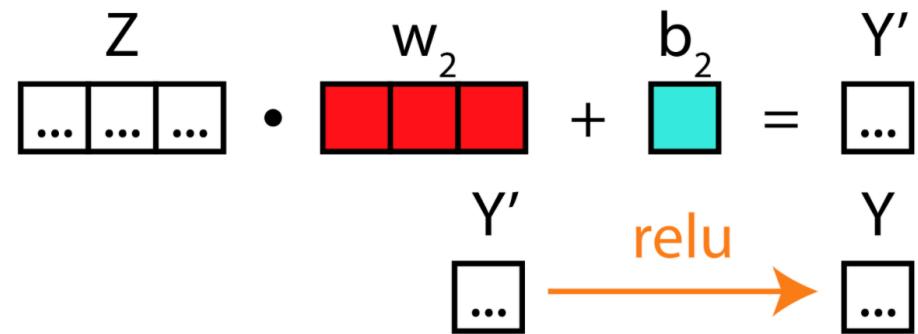
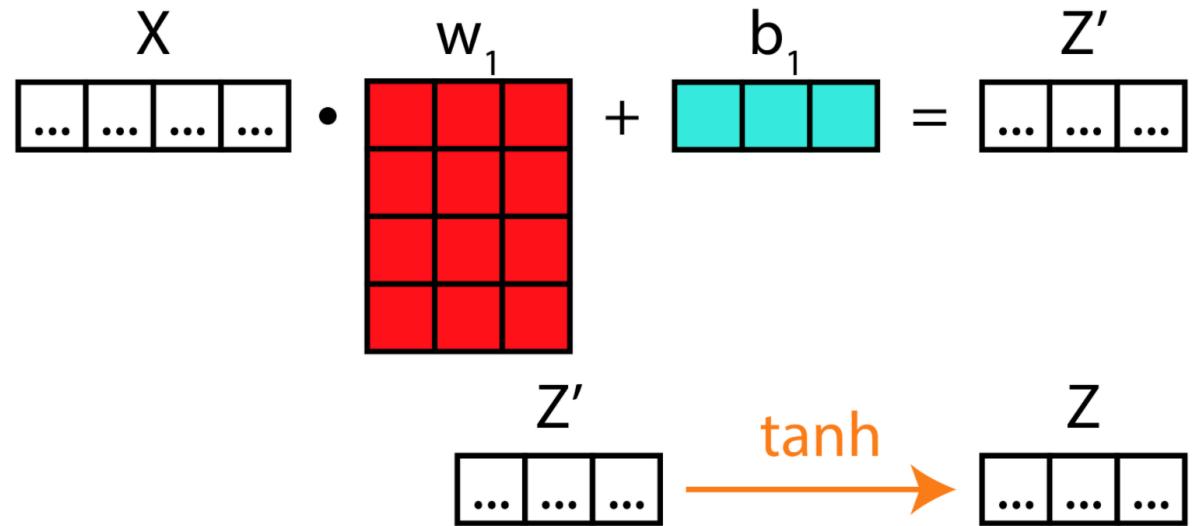
$$Y = \text{ReLU}(f(Z, w_2, b_2))$$

$$\text{ReLU}(x) = \max(x, 0)$$

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$$

Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

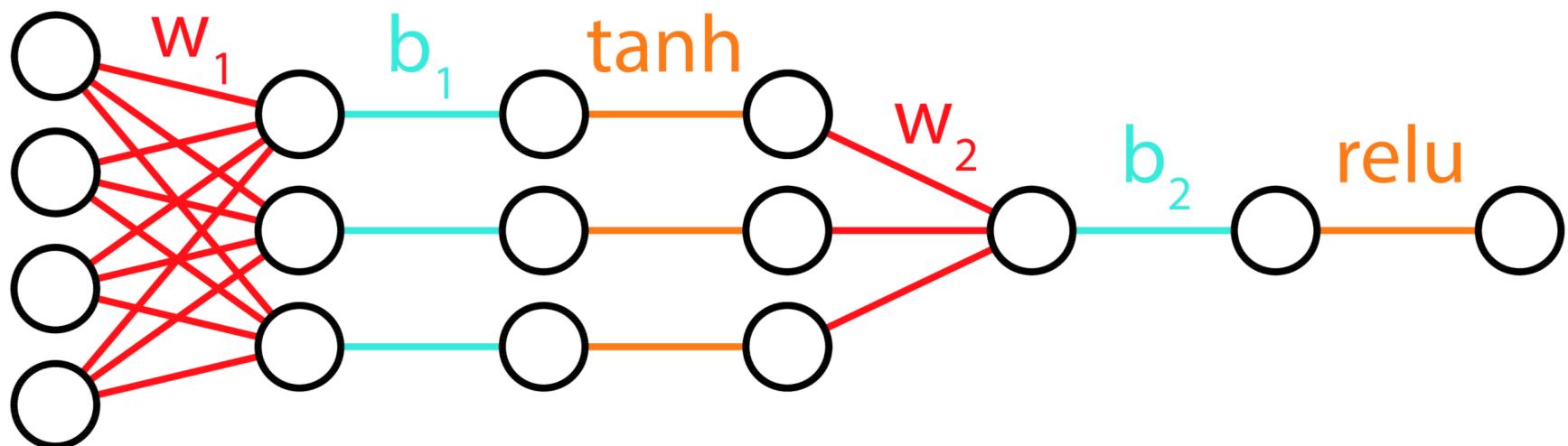
Matrices



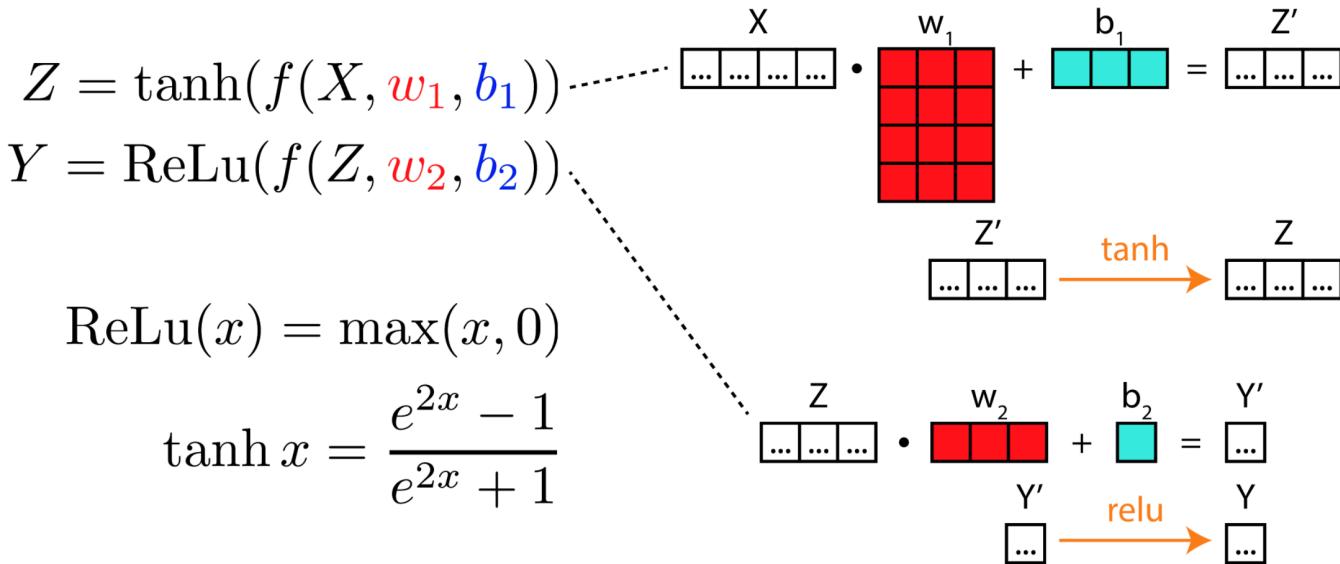
Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

Let's take a Neural View

Neural Diagram



DeepNets 3 Ways



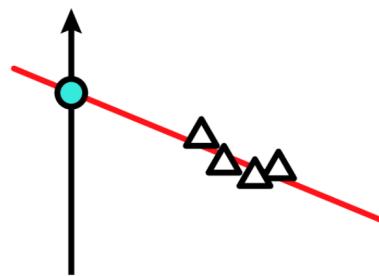
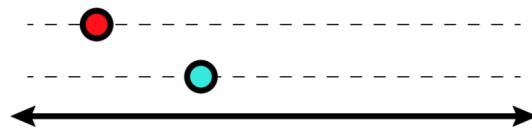
Source: [EricMjl/bayesian-deep-learning-demystified](https://ericmjl/bayesian-deep-learning-demystified)

Bayesian Neural Networks

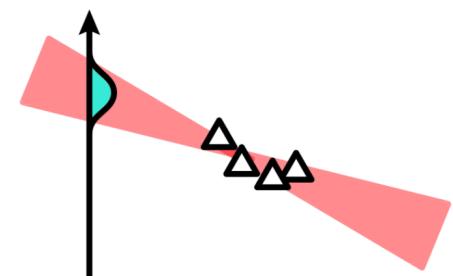
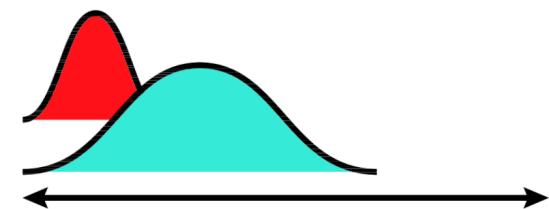
Key Idea: Learn probability density over parameter space.

Intuition

Non-Bayesian



Bayesian



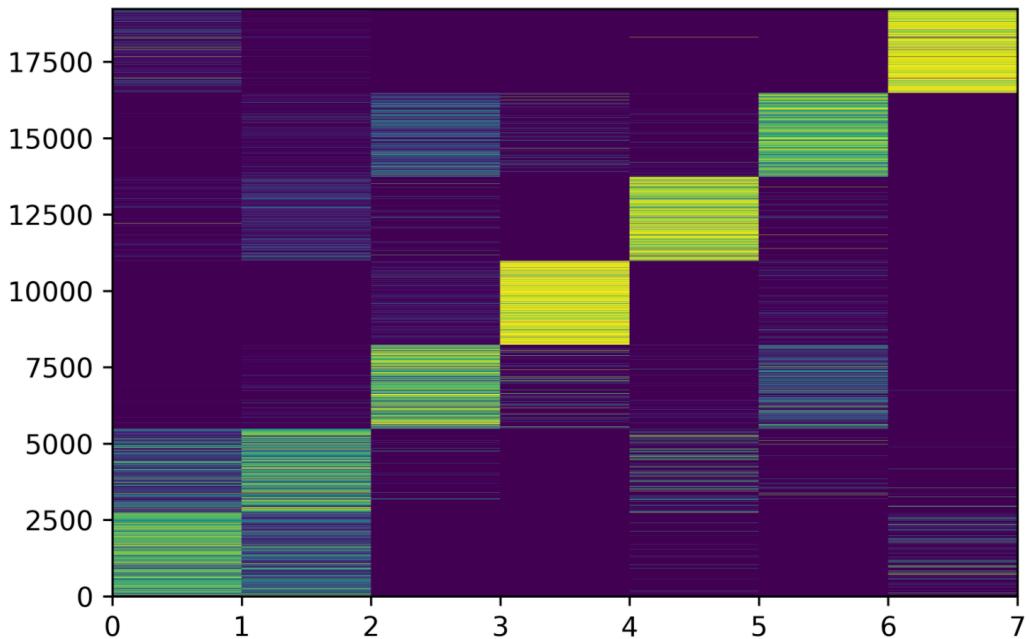
Source: [ericmjl/bayesian-deep-learning-demystified](https://ericmjl.github.io/bayesian-deep-learning-demystified)

Cheat Sheet

	function	matrices	computational graph	priors	postiors
linear regression	$Y = f(X, \mathbf{w}_1, \mathbf{b}_1)$ $f(x, w, b) = x \cdot w + b$			$w_{1,i} \sim N(0, 100)$ $b_1 \sim N(0, 100)$	
logistic regression (classification)	$Z = f(X, \mathbf{w}_1, \mathbf{b}_1)$ $Y = g(Z)$ $g(x) = \frac{1}{1 + e^{-x}}$			$w_{1,i} \sim N(0, 100)$ $b_1 \sim N(0, 100)$	
deep net regressor	$Z = \tanh(f(X, \mathbf{w}_1, \mathbf{b}_1))$ $Y = \text{ReLU}(f(Z, \mathbf{w}_2, \mathbf{b}_2))$ $\text{ReLU}(x) = \max(x, 0)$ $\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$			$w_{1,i} \sim N(0, 1)$ $b_{1,i} \sim N(0, 1)$ $w_{2,i} \sim N(0, 1)$ $b_{2,i} \sim N(0, 1)$	

Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

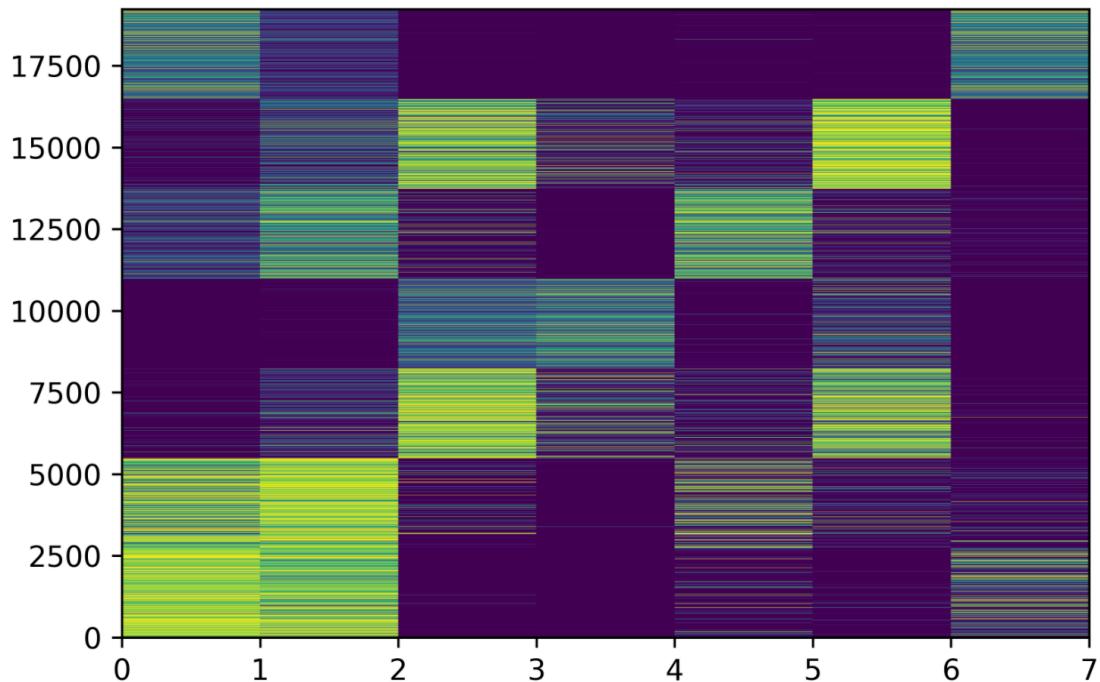
Predict Forest Cover Type Class Probabilities



"probabilistic estimate"

Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

Predict Forest Cover Type Class Uncertainties



"with uncertainties!"

Source: [ericmjl/bayesian-deep-learning-demystified](https://github.com/ericmjl/bayesian-deep-learning-demystified)

Take-Home Point 1

Deep Learning is nothing more than **compositions of functions on matrices**.

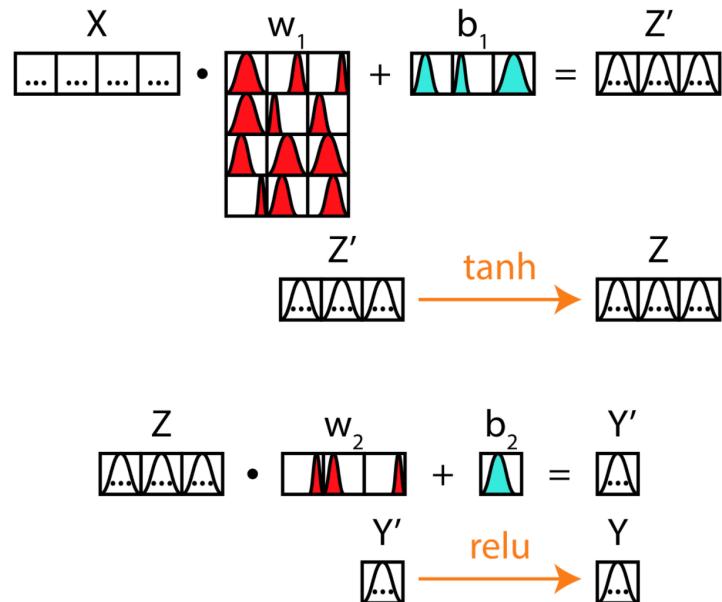
$$\begin{matrix} X \\ \cdots \cdots \cdots \cdots \end{matrix} \cdot \begin{matrix} w_1 \\ \text{red grid} \end{matrix} + \begin{matrix} b_1 \\ \text{cyan vector} \end{matrix} = \begin{matrix} Z' \\ \cdots \cdots \cdots \end{matrix}$$
$$Z' \xrightarrow{\text{tanh}} Z$$

$$\begin{matrix} Z \\ \cdots \cdots \cdots \end{matrix} \cdot \begin{matrix} w_2 \\ \text{red grid} \end{matrix} + \begin{matrix} b_2 \\ \text{cyan vector} \end{matrix} = \begin{matrix} Y' \\ \square \cdots \end{matrix}$$
$$Y' \xrightarrow{\text{relu}} Y$$

Source: ericmjl/bayesian-deep-learning-demystified

Take-Home Point 2

Bayesian deep learning is grounded on **learning a probability distribution for each parameter.**



Source: ericmjl/bayesian-deep-learning-demystified



Location as a latent variable / latent manifold in the prediction of house prices

Num tenants	age
	50

rooms	Sq footage	bathrooms	Address (zipcode)		Computed location index value
3	1500	1.5	12345		

$$\text{Rmse} = r1 \gg 0$$

Proximity to schools	Prox to shopping	Prox to highway	Prox to rest	walkability	zipcode
					12345

Gini/ f imp

$$\text{Loc index for buyers} = .2s + .4sh + .2h + .2r$$

$$\text{Loc index for renters} = .001s + .4sh + .2h + .05r + .7w$$



Hidden variables



Latent Space

Use-case : Stock Clustering

- Objective function → cluster types (understand the significance of the clusters; the meaning behind each cluster each segment)
 - Highest gain clusters
 - Consider other functions (objective) , label the new clusters
 - One dataset 2 additional columns one for each cluster type
 - Highest annual revenue, Lowest variance, $f(x) + g(x) = h(x)$
 - cluster on the one composite function
- Standards for labeling
 - Categorical , integer (1,2,3,) , top, middle, lower → numbers (categorical mapping) CountVectorizer in scikit learn

In class Group task: 6:36 – 6:55 pm

- Huddle on Latent Variable takeaways from Chopra Paper
- Re-address your project
- Be prepared to present it
- And talk about your fractal clustering

Chopra Paper Takeaways

- Ensemble model
- Latent manifold
- Data Amalgamation
- Read abstract, intro, conclusion, then scan titles and look for eg latent variable section.



A large, stylized text graphic is centered over a background of abstract, blurred blue and white diagonal stripes. The text reads "deep context AI" in a bold, three-dimensional font. The letters are primarily black with a blue outline, except for the "A" which has a red outline. The "deep" and "context" parts are stacked vertically, while "AI" is positioned below them. A faint watermark of a person's face is visible in the center of the background.

deep
context
AI